Recombinant process for preparing a complete malaria antigen, gp190/MSP1

The invention concerns a recombinant manufacturing process for the complete malaria antigen gp190/MSP1, as well as separate naturally-occurring domains and parts of the same, by expression of a synthetic DNA sequence. The invention concerns in addition the DNA sequences produced by the process and the host organisms used for the expression of the DNA sequences. In addition the invention concerns the use of the complete malaria antigen as well as parts thereof as a vaccine for immunization against malaria.

Finally the invention under consideration concerns a stabilization process for AT-rich genes, as well as stabilized genes which are characterized by a reduced AT content.

Malaria is one of the most significant infectious diseases in the world. According to WHO reports, in 1990 40% of the world population in 99 countries was exposed to the risk of malaria. At the same time its distribution is enormously on the increase. This may be ascribed above all to intensive development of resistance in the parasites causing malaria, promoted by the recommendation and use as prophylactics of the drugs intended for treatment. Besides the search for new and effective chemotherapeutic agents hope is nowadays directed towards the development of vaccines, since people in areas of the world where malaria is epidemic do manage to develop some kinds of immunity. As well as a natural resistance to malaria, such as that found in heterozygous carriers of the sickle-cell gene and people with thalassaemia and glucose-6-phosphate dehydrogenase deficiency, in the course of malarial infection in humans immune mechanisms can be stimulated which express themselves in a heightened capacity for resistance to the Plasmodia. Consequently the course of the disease in populations exposed to severe epidemics is generally less threatening than in persons exposed to the infection less frequently or for the first time.

The main problem in the development of a vaccine is the identification of an antigen which can induce protective immunity, since there is no easily accessible well-defined animal model available for the four parasites affecting man. The organism causing malaria belongs to the Plasmodium group, of which infection with one of the four parasites Plasmodium vivax,

Plasmodium ovale, Plasmodium malariae or Plasmodium falciparum results from the bite of Anopheles mosquitoes. Of these parasites Plasmodium falciparum is the most dangerous and the most widely distributed.

The main surface protein of the merozoite, the invasive form of the blood stage of the malaria parasite Plasmodium falciparum and other malaria parasites such as P. vivax, is a 190-220 kD glycoprotein. Late in the development of the parasite this precursor is processed into smaller proteins, which can however be isolated from merozoites as a unitary complex. By means of a glycosylphosphatidyl-inositol bond this complex is coupled to the merozoite membrane. The sequences of the gp190 proteins of various P. falciparum strains fall into two groups, between which intragenic recombination is frequent. In general the protein consists of many highly conserved regions, of a dimorphic zone to which in each case one of two alleles belongs, and of two relatively small oligomorphic blocs in the N-terminal zone (Tanabe, K., Mackay, M., Goman, M. and Scaife, J.G. (1987), Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum. J. Mol. Biol. 195, 273-287, Miller, L.H., Roberts, T., Shahabuddin, M. and McCutchan, T.F. (1993), Analysis of sequence diversity in the Plasmodium falciparum merozoite surface protein-1 (MSP-1). Mol. Biochem. Parasitol. 59, 1-14).

Already early on gp190/MSP1 was considered as a possible candidate for a vaccine. In the rodent model active protection against infection with rodent parasites was obtained following immunization with the analogous protein. Passive protection could be procured with antibodies directed against this protein (see also Holder, A.A. and Freeman, R.R. (1981), Immunization against blood-stage rodent malaria using purified parasite antigens, Nature 294, 361-364; Marjarian, W.R., Daly, T.M., Weidanz, W.P. and Long, C.A. (1984), Passive immunization against murine malaria with an IgG3 monoclonal antibody, J. Immunol. 132, 3131-3137). The data which ought to support this assumption are nevertheless in details not statistically significant.

There are in addition, a number of monoclonal antibodies which in vitro inhibit the invasion of erythrocytes by P. falciparum and are directed against gp190/MSP1 (Pirson, P.J. and Perkins, M.E. (1985), Characterization with monoclonal antibodies of a surface antigen of Plasmodium

falciparum merozoites. J. Immunol. 134, 1946-1951; Blackman, M.J., Heidrich, H.-G., Donachie, S., McBride, J.S. and Holder, A.A. (1990), A single fragment of a malaria merozoite surface protein remains on the parasite during red-cell invasion and is the target of invasion-inhibiting antibodies. J. Exp. Med. 172, 379-382).

Finally, a series of vaccine studies have been carried out with gp190/MSP1 from P. falciparum on primates, particularly on Aotus and Saimiri monkeys (see also Perrin, L.H., Merkli, B., Loche, M., Chizzolini, C., Smart, J. and Richle, R. (1984), Antimalarial immunity in Saimiri monkeys. Immunization with surface components of asexual blood stages, J. Exp. Med. 160, 441-451; Hall, R., Hyde, J.E., Goman, M., Simmons, D.L., Hope, I.A., Mackay, M. and Scaife, J.G. (1984), Major surface antigen gene of a human malaria parasite cloned and expressed in bacteria, Nature 311, 379-382; Siddiqui, W.A., Tam, L.Q., Kramer, K.J., Hui, G.S.N., Case, S.E., Yamaga, K.M., Chang, S.P., Chan, E.B.T. and Kan, S.-C. (1987), Merozoite surface coat precursor protein completely protects Aotus monkeys against Plasmodium falciparum malaria, Proc. Natl. Acad. Sci. USA 84, 3014-3018; Ettlinger, H.M., Caspers, P., Materile, H., Schoenfeld H.-J., Stueber, D. and Takacs, B. (1991), Ability of recombinant or native proteins to protect monkeys against heterologous challenge with Plasmodium falciparum, Inf. Imm. 59, 3498-3503; Holder, A.A., Freeman, R.R. and Nicholls, S.C. (1988), Immunization against Plasmodium falciparum with recombinant polypeptides produced in Escherichia coli, Parasite Immunol. 10, 607-617; Herrera, S., Herrera, M.A., Perlaza, B.L., Burki, Y., Caspers, P., Doebeli, H., Rotmann D. and Certa, U. (1990), Immunization of Aotus monkeys with Plasmodium falciparum blood-stage recombinant proteins, Proc. Natl, Acad. Sci. USA 87, 4017-4021; Herrera, M.A., Rosero, F., Herrera, S., Caspers, P., Rotmann, D., Sinigaglia, F. and Certa, U. (1992), Protection against malaria in Aotus monkeys immunized with a recombinant blood-stage antigen fused to a universal T-cell epitope; correlation of serum gamma interferon levels with protection, Inf. Imm. 60, 154-158; Patarroyo, M.E., Romero, P., Torres, M.L., Clavijo, P., Moreno, A., Martinez A., Rodriquez, R., Guzmann, F. and Cabezas, E. (1987), Induction of protective immunity against experimental infection with malaria using synthetic peptides, Nature 328, 629-632). In these vaccine studies two premises may be distinguished:

- Use of material isolated from parasites, and

- Administration of material procured in heterologous systems of expression.

The latter consists as a rule of relatively small segments of the total protein. Although the results of the inoculations carried out preliminarily on monkeys indicate that gp190/MSP1 could bring about protection, all the experiments carried out on primates have two problems, which place such a conclusion in question:

(a) they were carried out on too small groups of animals
(b) they were not repeated.

The results and the conclusions drawn from them are consequently not statistically confirmed. Besides, the difficulty of access to suitable monkeys there remains the main basic problem, that it has so far not been possible to manufacture good vaccination material in a suitable quantity.

On the other hand, after the sequencing of the gp190 gene from the K1 and MAD20 strains of Plasmodium falciparum overlapping fragments could be expressed in E. coli. With this material epidemiological studies in West Africa showed that in the adolescent group a correlation existed between antibody titre against gp190/MSP1 fragments on one hand and protection from parasite infection on the other. In addition the titre also appeared to correlate with the capacity to control the parasitaemia even at a low level (Tolle et al. (1993): A prospective study of the association between the human humoral immune response to Plasmodium falciparum blood stage antigen gp190 and control of malarial infections, Infect. Immun. 61, 40-47). These results are supplemented by new investigations on Aotus monkeys in the framework of the present invention. Here an enhanced protection against infection with the parasite was attained because protein preparations from Plasmodium falciparum, which consisted predominantly of unprocessed gp190/MSP1, had been used as vaccine. The monkeys with the highest antibody titres against gp190/MSP1 were the best protected. These results eventually indicated gp190 as a most promising candidate for a vaccine against tropical malaria.

By some groups of workers the C-terminal domain of gp190 (p19 or p42) is assigned a particular role in the immunity mediated by gp190 (see also Chang, S.P., Case, S.E., Gosnell, W.L., Hashimoto, A., Kramer, K.J., Tam, L.Q., Hashiro, C.Q., Nikaido, C.M., Gibson, H.L., Lee-Ng, C.T., Barr, P.J., Yokota, B.T. and Hui, G.S.N. (1996), A recombinant baculovirus 42-kilodalton C-terminal fragment of Plasmodium falciparum merozoite surface protein 1 protects Aotus monkeys against malaria, Inf. Imm. 64, 253-261; Burghaus, P.A., Wellde, B.T., Hall, T., Richards, R.L., Egan, A.F., Riley, E.M., Ripley-Ballou, W. and Holder A.A. (1996), Immunization of Aotus nancymai with recombinant C-terminus of Plasmodium falciparum merozoite surface protein 1 in liposomes and alum adjuvant does not induce protection against a challenge infection, Inf. Imm., in press.

Thus far, however, it has also been impossible to exclude other parts of gp190 on a rational basis as irrelevant to a protective immune response. Hence it is as necessary as ever to use the entire gene or the intact gp190 for vaccine investigations. Despite multiple investigations by various work-groups, however, there has not yet been any success in cloning and expressing the entire gp190/MSP1 gene.

Nor has it so far been possible to exclude a priori any part of the gp190 sequence as irrelevant to the protective immune response, so that it is as necessary as ever to use the entire gene or gene product for vaccine investigations. Nevertheless, despite many investigations by a number of working groups there has not yet been any successful cloning of the whole gene for gp190/MSP1.

One object of the present invention has consequently been to make available an adequate quantity of vaccine material in the form of the complete gp190/MSP1. It was a further object of the present invention to provide a process by which this vaccine material could be recovered.

In addition it was another object on the part of the present invention to provide a complete DNA sequence of gp190/MSP1 which could be expressed in a host organism.

Yet another object of the present invention was to provide host organisms containing the complete gp190/MSP1 gene.

Finally, it was also an object of the present invention to provide a stabilization process for AT-rich genes, as well as a stabilized gene suitable for expression characterized in a reduction of the AT content.

These objects are solved by the subject matter outlined in the Claims.

In the following, certain concepts are explained in more detail in order to make clear how they should be understood in this context.

"Recombinant manufacturing process" means that a protein of a DNA sequence is expressed by a suitable host organism in which the DNA sequence has arisen from cloning and fusion of individual DNA fragments.

"Complete gp190/MSP1 protein" here means the entire gp190/MSP1 surface protein isolatable from the above named Plasmodia, especially Plasmodium falciparum, representing the main surface protein of the above named parasite as well as the proteins with analogous function from the other Plasmodium species such as P. vivax. The term therefore comprises in each case the main surface protein of the merozoites of the four malaria parasites named above as dangerous to man. "Complete gp190/MSP1 gene" means the gene coding for this protein. In this context "complete" signifies that the entire amino-acid sequence of the native protein is present or that the gene sequence codes for the entire amino-acid sequence of the native protein. Mutated and/or shortened forms of gp190/MSP1 are however included therewith insofar as they display the same immunization potential (vaccine protection) as the complete gp190/MSP1. Finally the term also includes variants of gp190/MSP1 characterized by containing in one molecule protein fragments of various alleles.

"FCB-1" is a strain of P. falciparum such as that described in Heidrich, H.-G., Miettinen-Baumann, A., Eckerskorn, C. and Lottspeich, F. (1989) The N-terminal amino acid sequences of the Plasmodium falciparum (FCB1) merozoite surface antigens of 42 and 36 kilodalton, both derived from the 185-195-kilodalton precursor. Mol. Biochem. Parasitol. 34, 147-154.

"Attachment signal" here means a protein structure coded for by a DNA sequence at the 3' or 5' end of the gene according to the invention. Attachment signals are structures enabling the attachment of a polypeptide to other structures, such as for example membranes.

"Signal peptide" here signifies a protein structure for which a DNA sequence at the N-terminal end of the gene according to the invention codes. Signal peptides are structures which among other things enable penetration of the polypeptide into membranes.

In the context of the present invention "AT-content" means the percentage amount of adenine-thymine base pairs compared to guanine-cytosine base pairs.

"Cloning" will comprehend here all known state-of-the-art cloning methods which could be applied here, which are nevertheless not all described in detail because they belong among the normal tools of the person skilled in the art.

"Expression in an appropriate expression system" should here include all known state-of-the-art methods of expression in known expression systems which could be applied here, but which are nevertheless not all described in detail because they belong among the normal tools of the person skilled in the art.

It is a primary object in regard to the present invention that a process be provided by which the protein pg190/MSP1 and its gene can be produced in sufficient quantity without excessive cost.

This object is solved by the recombinant manufacturing process set out in Claim 1, by which a complete gp190/MSP1 gene and the protein coded by it are obtainable in sufficient quantities.

For the first time it is possible by this process to synthesize the protein in its entirety outside the parasite. As the analysis with conformational epitope-recognizing monoclonal antibodies shows, the protein thus synthesized is at least reproducibly synthesizable over wide areas in naturally folded form. By the recombinant manufacturing process many milligrams of intact gp190/MSP1 could in every case be recovered from the host organism, a quantity which for

technical and economic reasons can never be recovered from parasites. Production of the protein in any desired quantity is now possible and opens new perspectives for its use as an experimental vaccine against malaria. Furthermore, the way is now open for the development of living vaccines as well as for vaccines based on nucleic acids.

Synthesis of the gene sequence coding for the protein gp190/MSP1 is preferentially based on the sequence of the FCB-1 strain of P. falciparum. P. falciparum is the agent of tropical malaria and hence of the most dangerous among the types of malaria. The basic gene is a representative of the "K1 allele", where K1 stands for a particular P. falciparum strain. Its coding sequence extends over 4917 base pairs and includes a signal sequence at the N-terminal end as well as an attachment sequence at the C-terminal end.

Furthermore, according to the invention the recombinant manufacturing process is preferentially characterized in having the AT content of the DNA sequence on which the protein is based reduced relative to the wild type, from 74% in the original gene preferably to about 55%, for example while the amino-acid sequence of the FCB-1 protein is maintained a DNA sequence with the codon frequencies usual in the human genome is produced. Other codon frequencies which reduce the AT content are also conceivable.

Preferentially the gene underlying the protein produced by the recombinant manufacturing process codes for the full amino-acid sequence including signal peptide and GPI attachment signal peptide, further described as $gp190^{S}$.

In another preferred embodiment, the gene on which the protein produced by the recombinant manufacturing process is based codes for the complete amino-acid sequence except for the GPI attachment signal. This embodiment is then described as $gp190^{S1}$.

In yet another preferred embodiment, the gene on which the protein produced by the recombinant manufacturing process is based codes for the complete amino-acid sequence except for the GPI attachment signal and the signal peptide. This embodiment is then described as $gp190^{S2}$.

In a further preferred embodiment type, the gene on which the protein produced by the recombinant manufacturing process is based codes for the complete amino-acid sequence and a trans-membrane attachment sequence.

In a particularly preferred embodiment the recombinant manufacturing process includes the following steps:

In the first place the design of the DNA sequence to be synthesized on the basis of the gene from P. falciparum FCB-1, in which a DNA sequence with for example the codon frequencies common in the human genome is manufactured with retention of the amino-acid sequence of the FCB-1 protein.

The AT content of the gene should be reduced by this, preferably to 55%. Further on in the process the planned sequence is divided for example into 5 overlapping regions, which at the same time correspond to domains of the natural processing products of gp190/MSP1 from FCB-1: p83, p31, p36, p30 and p19.

Desoxyoligonucleotides are synthesized, which in each case extend the entire length of a region.

The desoxyoligonucleotides so synthesized are particularly preferred where their sequence corresponds in an alternating manner to the "upper" (5' - 3') or the "lower" (3' - 5') DNA strand. The length of these oligonucleotides is preferably on average 120 nucleotides and they overlap the neighboring sequences in each case by about 20 bases.

In one possible embodiment DNA sequences of about double the length of the existing end-products are manufactured by asymmetrical PCR, in effect so that the superfluous DNA sequences nearby in each case represent the opposite strand. This leads in a second

PCR amplification cycle to a second product corresponding to the length of four originally inserted oligonucleotides excluding the overlapping region. Transfer of these products to a preparation consisting predominantly of single-stranded DNA by asymmetrical PCR with the

terminal oligonucleotides permits the manufacture in a further amplification step of an 800-bp long double-stranded DNA fragment in only 25 PCR cycles.

In this manner the regions coding for p19, p30, p36 and p31 are directly synthesized and molecularly cloned in E. coli. Clones with fault-free sequences are conserved either directly or by the joining up of fault-free sequence fragments. The region which codes for p83 is constructed by fusion from two sequences comprising about 1200 bp.

In the further course of production single sequences are cloned. As expression vectors candidates preferred are the plasmids pDS56, RBS11 ("Hochuli, E., Bannwarth, W., Doebeli, H., Gentz, R. and Stueber, D. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. Biotechn. 6, 1321-1325"), pBi-5 ("Baron, U., Freundlib, S., Gossen, M. and Bujard, H. (1995) Corregulation of two gene activities by tetracycline via a bidirectional promoter. Nucl. Acids Res. 23, 3605-3606") and ppTMCS. It is possible nonetheless also to conceive of other expression vectors.

Host organisms preferred for expression are E.coli, with the strain DH5alphaZ1 especially preferred (R. Rutz, Dissertation 1996, Heidelberg University), HeLa cells, CHO cells, Toxoplasma gondii (Pfefferkorn, E.R. and Pfefferkorn, C.C. 1976, Toxoplasma gondii:Isolation and preliminary characterization of temperature-sensitive mutants. Exp. Parasitol. 39, 365-376) or Leishmania. Additional host systems might be e.g. yeasts, baculoviruses or adenoviruses, so that the subject matter of the invention should not be limited to the host systems mentioned.

A further object of the present invention has been to provide a complete DNA sequence, suitable for expression, of the gp190/MSP1 surface protein of P. falciparum.

This object is solved by the invention named in Claim 17, by which the sequence can be obtained by the recombinant manufacturing process described above.

In a preferred embodiment of the present invention the sequence suitable for expression codes for the complete amino-acid sequence.

In another preferred embodiment of the present invention the sequence suitable for expression codes for the complete amino-acid sequence except for the attachment signal.

In a further preferred embodiment according to the present invention the DNA sequence suitable for expression codes for the complete amino-acid sequence except the attachment signal and the peptide signal. This embodiment of gp190/MSP1 can hence be characterized in including at the N-terminus 11 additional amino-acids, of which 6 are histidines.

Particularly preferred the DNA sequence suitable for expression contains no recognizable "splice-donor" and "splice-acceptor" sites, and is preferably characterized in not containing any larger GC-rich sequences which might result in stable hairpin structures at the RNA level.

Recognition signals for restriction enzymes which recognize sequences of six or more base pairs should preferably be avoided.

In a preferred embodiment specific cleavage sites for restriction endonucleases, occurring only once in the gene, are introduced into regions to separate the existing domains following processing of the protein.

Particularly preferred would be the presence at both ends of the gene of sequences for restriction endonucleases which do not occur in the gene.

Furthermore host organisms containing the complete sequence of gp190/MSP1 surface protein are provided by the invention.

Such host organisms are preferably E. coli, particularly preferred being the strain DH5alphaZ1, HeLa cells, CHO cells, Toxoplasma gondii or Leishmania. The HeLa and CHO cells ought preferably to synthesize constitutively tTA.

Finally the present invention provides a possibility of using a gp190/MSP1 surface protein created produced according to the recombinant manufacturing process, or parts thereof, for active immunization against malaria.

The scheme for synthesis presented here also permits manufacture of the second allele of the gp190/MSP1 gene, whereby the dimorphism of the protein is also taken into account. The main variability of the protein depends however on the sequences of two relatively short blocs, blocks II and IV (ref. 1), which are oligomorphic. The present sequence data make it possible to disclose over 95% of all known gp190/MSP1 sequences with 6-8 sequence combinations of these blocs. The synthesis of these sequence variants can be brought about problem-free by means of the strategies proposed here, so that variants can be built up both in the K1 and in the MAD20 allele. Vaccines from the families of sequences thus created can confer protection where required against a wide spectrum of parasites with gp190/MSP1 variants.

The manufacture of different types of vaccine is possible:

- At the level of protein preparations, where in each instance mixtures of the two families (K1 type, MAD20 type with different variants of Blocs II and IV) can come into use. Various carrier or adjuvant materials could be added: aluminum oxide, liposomes, IscomsQSz1, etc.

- At the level of live vaccines: (a) viral carriers, especially vaccinia and adenoviruses; (b) parasites as carriers, particularly avirulent forms of Leishmania and Toxoplasma; (c) bacterial carriers, e.g. Salmonella.

- At the level of nucleic acids, whereby for example vectors suitable for gene therapy would be used to introduce the gene into the host; beyond that the introduction of nucleic acids coding for the desired protein can be envisaged.

A further possibility for vaccination lies in the use of a gp190/MSP1 protein produced according to the recombinant manufacturing process set out by the invention, for the production of monoclonal antibodies which can then be used in their turn for passive immunization against malaria.

Similarly it becomes possible to use the DNA sequence on which the protein is based at an intermediate stage arising in the course of the recombinant manufacturing process for the construction of a vaccine based on nucleic acids.

Finally the invention also concerns a process for the stabilization of gene sequences, especially for sequences which do not show adequate stability in expression systems.

According to the invention this stabilization is attained because the AT content of the sequence is reduced.

Moreover a stabilized gene characterized by having a reduced AT content is provided by the invention. An example of such a stabilized gene is the gene for gp190/MSP1 surface protein according to the present invention.

In the following the invention will be described with the help of figures and tables as well as some examples in individual embodiments.

They show:

Fig. 1: Schematic representation of the gp190/MSP1 precursor protein from P. falciparum (FCB-1).

Fig. 2: Two vaccine trials carried out on Aotus monkeys with native gp190/MSP1 from P. falciparum (FCB-1).

Fig. 2A: With 3 x 60 micrograms gp190/MSP1

Fig. 2B: With 3 x 40 micrograms gp190/MSP1

Fig. 3A: Strategy of synthesis of the gp190/MSP1 gene

Fig. 3B: Principle of PCR-based total synthesis

INS C1 Fig. 3C: Total sequencing of gp190^{S}

INS C2 Fig. 3D: N- and C-termini of gp190^{S1} variant

Fig. 4A: Expression vector pDS56 with gp190^{S2} sequence

Fig. 4B: Gel electrophoresis of gp190^{S2}

Fig. 5A: Expression vector pBi-5 with gp190^{S1} sequence

Fig. 5B: Immunofluorescence of HeLa cells

Fig. 5C: Electrophoretic characterization of gp190^{S1} purified from HeLa cells

Fig. 6A: Expression vector ppT 190 with gp190 sequence

Fig. 6B: Immunofluorescence of the expression of gp190^{S} in T. gondii

Fig. 6C: Polyacrylamide gel electrophoresis of gp190 from T. gondii

In the gp190/MSP1 precursor protein from P. falciparum schematically represented in Fig. 1 the dark blocs stand for regions which are strongly conserved in all strains. The cross-hatched blocs indicate the dimorphic areas, which in the case of the FCB-1 isolate derive from the K1 allele. O1 and O2 indicate the oligomorphic areas. S denotes the peptide signal sequence containing 19 amino-acids, GA the C-terminal region, which includes the signal for the GPI attachment of the protein to the membrane. The arrows indicate the sites of the processing by

which the proteins p53, p31, p36, p30 and p19 arise. The gp190 gene codes for altogether 1639 amino-acids.

The other figures are more conveniently explained in the context of the following Examples.

EXAMPLES

Example 1: total synthesis of one of the DNA sequences coding for gp190/MSP1 (see Fig. 3)

A. Strategy of synthesis of the gp190/MSP1 gene ($gp190^S$) (see Fig. 3A).

The sequence was divided into fragments corresponding to the main processing products: p83, p31, p36, p30 and p19. In the transition regions cleavage sites for restriction endonucleases (arrows in fig.3) were inserted in such a way that the amino-acid sequence was not altered. All the particular cleavage sites are found only once in the sequence.

The fragments were synthesized to overlap, so that the cleavage sites at the respective ends made attachment by ligation to the neighboring fragment possible. All individual fragments contain in addition at their 5' ends a BamHI division site for insertion into expression vectors. The entire sequence could be cloned via MluI and ClaI. The scheme indicated here leads in addition to a sequence which cannot produce the GPI attachment since the C-terminal lacks 18 amino-acids. Synthesis of a corresponding oligonucleotide as well as of a "primer" extending over the SphI cleavage site, leads after PCR to the GA fragment which could be used by SphI and ClaI, the resulting total sequence being $gp190^S$. On removing the sequence coding for the peptide signal, "PCR Primer" is produced, over which the fragment ΔS has been synthesized. It is permissible to alter the N-terminal via a BamHI and a HindIII cleavage site in such a way that the protein begins with amino-acid no. 20. The nuclear sequence which encodes gp190/MSP1 without signal sequence and without GPI attachment signal was designated gp190S2. Deletion of the GPI attachment signal alone leads to $gp190^{S1}$.

B. Principle of the PCR-supported total synthesis (see Fig. 3B)

Oligodesoxynucleotides of about 120 nucleotides have been synthesized in an alternating manner from the coding or the non-coding strand in such a way that in each case about 20 bases overlapped with the neighboring fragment. The scheme illustrates for example the synthesis of a fragment about 800bp long from oligonucleotides. At the first stage 2 oligonucleotides were amplified "asymmetrically" in each of 4 reaction vessels. This resulted in 4 populations of DNA about 220bp in length, consisting predominantly of single strands (A, B, C, D). Uniting A to B and C to D with amplification over 5 cycles led to 2 approximately 400bp long double-stranded products. Asymmetrical amplification of these DNA fragments (Stage III) resulted in single-stranded populations which following uniting and amplification (Stage IV) resulted after 10 cycles in the end-product G of about 800bp in length. This synthesis could be carried out without isolation of intermediate products and without renewing buffer or enzyme, and was completed in 3 hours. The end-product was purified electrophoretically, divided up with the appropriate restriction endonucleases, and cloned in E. coli in pBluescript (Stratagene), to which polylinker a MluI and a ClaI cleavage site had been added.

C. Total sequence of gp190^{s} (see Fig. 3C)

Following fusion of all part sequences (Fig. 3A) in pBluescript, the sequence of the gene was checked by the di-deoxy method. The reading frame of gp190^{S} had a length of 4917bp (+ 2 stop-codons) and encoded an amino-acid sequence corresponding to that of the gp190/MSP1 from FCB-1 (1639 amino-acids).

INS C3

D. N- and C-termini of the gp190^{S1} variant (see Fig. 3D)

INS C4

The N-terminal sequence, beginning with the BamHI cleavage site, indicates the transition at amino-acid 20, from which it can be assumed that after splitting of the signal peptide it defines the N-terminus. At the C-terminus the sequence encoded ended at amino-acid 1621. The stop-codon followed the ClaI cleavage site.

Example 2: Expression of gp190^{S2} in E. coli

A. Expression vector (see Fig. 4A)

The gp190^{S2} sequence was inserted via the BamHI and ClaI cleavage sites into pDS56RBSII, by means of which 6 histidines as well as some amino-acids originating in the vector were fused to the N-terminus. This produces the following N-terminal sequence on the reading-frame: Met Arg Gly Ser (His)$_6$ Gly Ser. Through the promoter $P_{N25lacO-1}$ the transcription comes under lacR/O/IPTG control.

INS C5

B. Expression and purification of gp190^{S2} (see Fig. 4D)

Carrying over the vector pDS56RBSIIgp190^{S2} into E. coli DH5alphaZ1 and induction of synthesis through IPTG resulted after electrophoretic separation of the total protein extract from the culture in a clearly visible band of the anticipated size (arrow). Purification of the material through IMAC and affinity chromatography (antibody column with mAK5.2) led to a homogeneous product of about 190 kD. In the Figure M stands for molecular weight standards; 1 = E. coli before; 2 = after induction with IPTG for 2 hours; 3, 4, 5 = fractions from elution of the mAK column.

Example. 3: Tetracycline-controlled expression of gp190^{S1} in HeLa and CHO cells and isolation of the product (see also Fig. 5 and 6c)

A. The gp190 sequence was inserted via the BamHI/ClaI cleavage sites into the expression vector pBi-5. In this way transcription of the gene came under the control of a bidirectional "tTA-reponsive" promoter and could be regulated through Tc. The bidirectional promoter simultaneously initiated transcription of the indicator gene luciferase. In consequence the regulation of the expression could easily be followed (see also Fig. 5A).

B. Immunofluorescence of HeLa cells, which express luciferase and gp190^{S1} under Tc control

The production of luciferase (left), gp190^{S1} (middle) in the absence of Tc was demonstrated in HtTA93-9 cells, which contain the bidirectional transcription unit of (A). Following addition of Tc no noteworthy synthesis of gp190S1 was shown (as represented in Fig. 5B,right).

C. Electrophoretic characterization of gp190^{S1} purified from HeLa cells

The HeLa cell clone HtTA93-9 as well as the CHO cell clone CHO27-29 have been cultivated with or without Tc. Cell extracts separated by electrophoresis have been analyzed with mAK5.2 by means of "Western blotting" (Fig. 5C); analysis of the CHO cell line is shown on the left, of HeLa on the right. (1) = culture without, (2) = culture with Tc, (3) = non-transfected HtTA-1 cell line. Molecular weight standards are in each case indicated on the left.

D. Purification of gp190^{S1} synthesised by HeLa cell clone HtTA93-9

Preparative cultivation of the HtTA line and induction of expression of gp190^{S1} by withholding Tc permitted isolation of the gene product by affinity chromatography (mAK5.2 column).

The polyacrylamide gel stained with Coomassie (Fig. 6C) following electrophoresis displayed a product consisting of gp190^{S1} as well as another protein of about 50 kD. The latter was not derived from gp190^{S1} and thus originated from the HeLa cells. Its projected removal should nevertheless present no difficulty in principle.

Example 4: Expression of gp190^{S1} in Toxoplasma gondii and purification of the product (see also Fig. 6).

A. The gp190^{S} sequence was inserted into the vector ppT via MluI/PstI. This brought the gene under the control of the tubulin promoter (P_{tub-1}) of T. gondii. The 3' untranslated region (VTR) originated from the main surface protein of T. gondii (SAG-1).

B. Expression of gp190^S in T. gondii

Transfection of T. gondii with pTT190 led to the isolation of parasite lines which expressed constitutively gp190^S. Immunofluorescence with mAK5.2 (middle picture) showed not only expression of the gene but also situated the binding of the expression product close to the surface of the parasite, since it, like SAG-1, provokes the same pattern of immunofluorescence (right section of fig. 6B). On the left in Fig. 6B a phase contrast photograph of the middle picture is shown.

C. Isolation of gp190S from T. gondii.

By means of affinity chromatography (mAK5.2 column) gp190^S was purified from a prepared quantity of T. gondii (5 x 10^9 parasites). The extremely pure protein possessed the anticipated molecular weight, as the Coomassie-stained polyacrylamide gel indicated following electrophoresis (2-3 on Figure 6C). At no. (1) on Fig. 6C purified gp190^{S1} from CHO cells is represented with molecular weight marked on the left side.

Example 5: Characterization of gp190^S with monoclonal antibodies.

The interaction of 16 monoclonal antibodies with gp190^S from the various heterologous expression systems was reviewed by immunofluorescence on P. falciparum and T. gondii or by "Western blot" on the purified proteins. Complete agreement was found when the two parasites were compared (number of +s indicates the relative intensity of the fluorescence). On Western blotting 12mAK's reacted with gp190^S from E. coli and T. gondii. On the other hand 3 antibodies did not bind to material isolated from CHO cells. Antibodies 15 and 16, which recognize epitopes from the oligomorphs or the alternative allele (MAD20), did not react. The results are summarized in Table 1, in which ND means "not carried out".

Example. 6: expression of gp190^{S} in heterologous systems

1. Expression in E. coli

The gp190^{S2} was inserted into the expression vector pDS56, RBSII, where it came under control of the promoter PN $_{25lacO-1}$, which can be controlled via the lac operator/repressor/IPTG system (Fig. 4A). Transfer of the plasmid into repressor-producing E. coli cells, eg E. coli DH5alphaZ1, permitted expression of pg190^{S2} under IPTG control. By means of a nickel-chelate column the raw product could be isolated via the N-terminal $(His)_6$ sequence introduced by the vector. An ensuing affinity chromatography on an antibody column led to an extremely pure preparation. Since the monoclonal antibodies used (mAK5.2) recognized a conformational epitope in the C-terminal region, this 2-step purification selected a full-length intact protein with correct folding at least at the C-terminus (Fig. 4B).

In contradistinction to the natural material the end-product possesses 11 additional amino-acids at the N-terminus, of which 6 are histidines. It contains no N-terminal signal and also no C-terminal attachment sequence. The P. falciparum-specific sequence begins with amino-acid 20 and ends with amino-acid 1621.

2. Controlled expression of gp190^{S1} in HeLa and CHO cell cultures

The gp190S1 was inserted into the vector pBi-5 and thereby placed under control of a promoter regulable by tetracycline (Tc). The Tc-contolled system was chosen for 2 reasons:

- It belongs to the expression systems with which the highest yield is obtained in mammalian cells.

- Unsecreted foreign proteins at high concentration can interfere negatively with cell metabolism. Synthesis of the desired product is consequently begun only after maturation of the culture.

In the construct pBi5-gp190^{S1} a bidirectional promoter was activated by the Tc-controlled transcription activator and initiated transcription of both gp190^{S1} and the luciferase indicator gene. In the presence of Tc the promoter is inactive. The transcription unit was transferred into both HeLa and CHO cells, which both synthesize constitutively tTA (HtTA line: Gossen, M. and Bujard, H. (1992), Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. USA 89, 5547-5551; CHO-tTA line, unpublished). Through cotransfection (Ca^{2+}-phosphate method) with a hygromycin-resistance-inducing marker gene was selected for successful chromosomal integration. Hygromycin-resistant clones were then investigated for regulability of the expression >Tc, in which luciferase activity was used as indicator. The gp190 synthesis was tested in well regulable clones (regulation factor •Tc 1000). Immunofluorescence analysis (Fig. 5B) as well as investigation by "Western blot" (Fig. 5C) allowed the identification in both cell types of clones which synthesized gp190 under strictly regulable conditions. The best regulable of 20 clones were in each case subcloned. The subclones HtTA93-9 and CHO27-29 were used for cultures on a scale of 10:1. From cell extracts of these cultures intact gp190^{S1} could be isolated by means of affinity chromatography (mAK5.2). The material was homogeneous except for a single cellular component which did not derive from gp190^{S1} and made up 25% of the preparation (Fig. 6C). It had to be removed in a further purification step.

3. Expression of gp190S in Toxoplasma gondii.

Like P. falciparum, Toxoplasma gondii belongs to the Apicomplexa and consequently has a protein modification system apparently similar to that of P. falciparum. T. gondii can be transfected with foreign DNA which is efficiently integrated into the genome and furthermore allows problem-free multiplication of T. gondii in cell culture. To obtain a product most like native gp190, gp190S2 is expressed in such a way that the protein is secreted on the surface of the parasite and, as in P. falciparum, bound to the membrane via a GPI analogue. In that way the gp190S2 (Fig. 3A) has been inserted (Fig. 6A) into the plasmid ppTMCS (D. Soldati, unpublished) and thereby placed under the control of the T. gondii tubulin promoter,

This expression construct was transfected into T. gondii. Selection with chloramphenicol led to resistant clones synthesizing gp190 which were detected by immunofluorescence (Fig. 6B).

The immunofluorescence with anti-gp190 antibodies was indistinguishable from a corresponding pigmentation of the parasites by means of antibodies against SAG1, the main surface protein of T. gondii. It may be deduced from this that gp190 is bound to the surface of T. gondii. Several T. gondii clones (Nos. 3.1 to 3.4) were characterized and saved for the production of gp190. Using affinity chromatography (mAK5.2) gp190 was isolated from T. gondii cultures (clone 3.4) cultivated on a preparative scale. Electrophoretic analysis revealed a homogeneous product with a migration rate similar to that of the intact protein (Fig. 6C)

Example 7: Characterization of gp190 protein from various expression systems by means of monoclonal antibodies.

A set of gp190-specific monoclonal antibodies, of which a number recognize conformational epitopes, were used to compare the reactivity of the antibodies with P. falciparum and T. gondii parasites via immunofluorescence. Table 1 shows that the reactivity of the 16 antibodies with either parasite is the same. This is a strong indication that in T. gondii "native" gp190 is being mostly produced. Comparison of the reactivity of the antibodies with protein from E.coli, HeLa or CHO cells as well as T gondii shows also that most of the antibodies react with the 4 preparations. In particular the protein derived from E. coli recognizes more of the antibodies than that produced in mammalian cells. This is apparently a consequence of glycosylation in mammalian cells.

Example 8: Immunization of Aotus lemurinus griseimembra monkeys with gp190/MSP from P. falciparum (FCB-1).

Two independent immunization experiments (A, B) were carried out. In them in one instance (A) 1.0 mg and in the other (B) 0.6 mg of very pure gp190/MSP1 was extracted from about 2 x 10^{11} parasites respectively.

The protein was administered together with Freund's Adjuvant (FCA). The control group received only FCA. Immunization equally with the protein mixture or the adjuvant was done three times at intervals of 4 weeks. Two weeks after the last immunization each of the animals

was infected with 10^5 parasites (FVO strain) from a donor animal. Parasitaemia was measured daily. The results are summarized in Fig. 2. The symbols mean:

T: that the animals were treated with resochin

D: a dead animal

Fig. 2A: individuals in the vaccinated group each received
3 x 60 micrograms gp190/MSP1

Fig. 2B: individuals in the vaccinated group each received
3 x 40 micrograms gp190/MSP1

While in the control group only 1/11 animals did not develop parasitaemia, this was 6/10 in the vaccinated group. The four animals in the vaccinated group who did develop a pronounced parasitaemia did so - in comparison to the control group - with an average delay of four days (exceeding the 2% limit of parasitaemia).

These experiments indicate for the first time a highly significant protection by gp190/MSP1 against infection with P. falciparum in a monkey model. The process according to the invention consequently permits a practical vaccine against malaria to be presented for the first time.

Table 1: Interaction of gp 190^S with monoclonal antibodies

					IFA		Western blot	
Code	mAb	Type of epitope	Variability	P.f. FCB	Toxoplasma	E. coli	Toxoplasma	CHO
1	5.2	conformational	conserved	++++	++++	+	+	+
2	12.10	conformational	conserved	++++	++++	+	+	+
3	7.5	conformational	conserved	++++	++++	+	+	+
4	12.8	conformational	conserved	++	++	+	+	+
5	7.3	conformational	dimorph (K1)	++++	+++	+	+	+
6	2.2	conformational	conserved	++++	++++	+	+	+
7	7.6	conformational	dimorph (K1)	++++	++++	+	+	+
8	9.8	conformational	conserved	++++	++	+	+	-
9	13.2	sequential	conserved	++++	++++	+	+	+
10	13.1	sequential	dimorph (K1)	++++	+++	+	+	-
11	6.1	sequential	dimorph (K1)	++++	++++	+	+	ND
12	A5Z	unknown	unknown	+++	+++	+	+	+
13	17.2	unknown	unknown	++++	+++	ND	ND	ND
14	15.2	unknown	unknown	++++	+++	ND	ND	ND
15	9.7	conformational	dimorph (MAD20)	-	-	-	-	-
16	12.1	sequential	oligomorph	-	-	-	-	-